



ELSEVIER

Available at

www.ElsevierComputerScience.com

POWERED BY SCIENCE @ DIRECT®

Computers & Graphics 28 (2004) 113–117

COMPUTERS
& GRAPHICS

www.elsevier.com/locate/cag

Chaos and graphics

A method for approximating missing data in spatial patterns

J.C. Sprott*

Department of Physics, University of Wisconsin, 1150 University Avenue, Madison, WI 53706, USA

Abstract

Spatial patterns such as historical landscape records or digital photographs are often plagued by large numbers of missing or otherwise corrupted data points or pixels that cannot be easily reproduced. A method is described in which a simple stochastic cellular automaton is used to produce fictitious fractal data at arbitrarily many spatial points such that the resulting pattern mimics the morphological features of the actual pattern. The method is simple to implement, preserves all the existing data, has no adjustable parameters, and can be used to fill in regions of arbitrary size and shape, even outside the region for which data are available. Furthermore, it reduces to more conventional interpolation methods when only a few isolated data points are missing.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Image processing; Cellular automata; Fractals

1. Introduction

Spatial patterns such as landscape data [1] and digital photographs are often plagued by missing, ambiguous, or otherwise corrupted points that can greatly hinder some types of analysis. In many cases, the data cannot be recovered because the pattern has since changed or is no longer accessible. There are many plausible schemes for guessing the identity of such points, for example, using logistic regression [2] or by assuming that the missing points are the same as their nearest neighbors or the same as the most abundant data type within some radius. However, these methods perform poorly with large blocks of missing data since they tend to fill in the blocks homogeneously, and they can greatly distort the morphological features of the data.

Digital image processing and enhancement have a vast and rich history [3–12] including methods involving cellular automata [13–15]. However, most traditional methods do not explicitly use the knowledge of which pixels are corrupted and typically smooth or sharpen the image uniformly, altering the good as well as the bad pixels in the process. These methods often use sophis-

ticated mathematics, not easily understood by students and programmers, have adjustable parameters, and are difficult to implement. Furthermore, traditional methods typically do not preserve the small-scale fractal structure within the large blocks of missing data, resulting in a blurry image.

2. The voter model

Recently, Bolliger et al. [16] proposed a stochastic cellular automaton model that produces a landscape resembling that from historical records starting from a wide variety of arbitrary initial conditions. The model is extremely simple and is known in the cellular automaton literature as the voter model [17] since it models a particularly impressionable political electorate. It consists of a two-dimensional rectangular grid of cells with each cell assuming one of a finite number of values that could correspond to landscape type (forest, savanna, prairie, etc.), tree type (pine, oak, maple, etc.), or any other characteristic such as predominant political regions of democrats and republicans. The model is initialized either randomly or with a highly ordered distribution of values having the same probability as the landscape to be modeled. At each iteration, cells are replaced by the contents of a cell chosen at random from

*Tel.: +1-608-263-4449; fax: +1-608-262-7205.

E-mail address: sprrott@physics.wisc.edu (J.C. Sprott).

within a circular neighborhood of radius r . The cells can be updated either synchronously (all at once) or asynchronously (randomly). Boundary conditions can either be periodic (the pattern wraps both horizontally and vertically), or reflecting (as with a mirror at the boundary).

After many iterations with an appropriately chosen value of r , the pattern evolves to one whose features strongly resemble the historical data according to various metrics. Common metrics are the cluster probability (the probability that an arbitrary cell is the same as its four nearest neighbors), fractal dimension, or algorithmic complexity [18]. The fractal pattern continues to evolve with scale-invariant temporal fluctuations having a $1/f^\alpha$ power spectrum with $\alpha \approx 1.6$ as would be expected for a self-organized critical system [19].

The method proposed here uses the cellular automaton model locally to fill in regions of missing or corrupted data without disturbing the remaining good data. The single parameter r is chosen so that the regions of missing data have the same cluster probability as the regions of good data, although any other appropriate metric could be substituted. The method can be used to supply a single missing point, which becomes the same as one of its randomly chosen neighbors within a radius r , or to generate an entire landscape as was done by Bolliger et al. [16], and thus it has wide applicability. The method is illustrated using a hypothetical two-component square data set with a large block of data missing from its interior. The extension to more complicated images with many more components is straightforward and is illustrated with digital photographs from which numerous large blocks of pixels are removed. Although the method is illustrated with raster data, it can be applied equally well to vector data in which the points do not lie on a regular grid.

3. Application to landscape data

On the left of Fig. 1 is a hypothetical two-component landscape generated by a stochastic cellular automaton with 208×208 cells after 1000 iterations, starting with equally probable but random values of 0 or 1 in each cell. Cells asynchronously (randomly) replaced by one of their eight nearest neighbors are also chosen randomly. This plot could represent, for example, regions of forested and non-forested landscapes, but it could also be a plot or digital photograph of anything whose underlying structure is a random fractal, such as clouds or regions of lakes and land. For this case, the probability that a cell is the same as its four nearest neighbors is about 40%. If this were an observed landscape with missing data and an irregular boundary, the cluster probability could be calculated for only those points in the set that have four correctly identified neighbors. Missing data and data outside the boundaries are presumed to have the same cluster probability.

Fig. 2 shows how the cluster probability changes with the size of the neighborhood from which replacements are chosen. Small neighborhoods exhibit strong organization with relatively few large clusters and a large cluster probability, while large neighborhoods exhibit weak organization with many small clusters and a small cluster probability. The data points represent the mean of 25 different realizations of the original data, and the error bars represent plus or minus one standard deviation. Note that a random two-component pattern with equal densities has a cluster probability of $(\frac{1}{2})^4 = 6.25\%$.

The parameter r , or equivalently the number of neighbors (approximately πr^2), can be chosen to mimic a wide range of patterns, including one with many missing data points. The replacements need not be chosen uniformly from a disk of radius r , but could, for

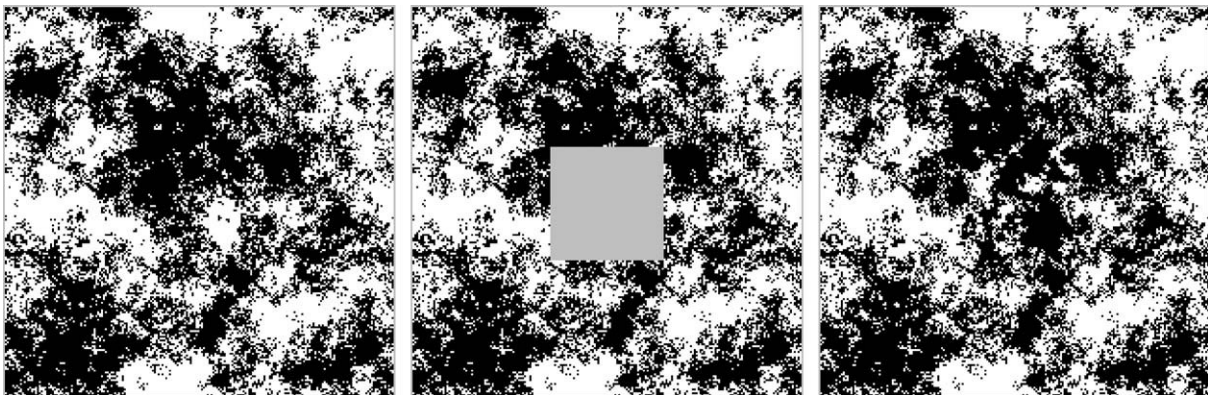


Fig. 1. Hypothetical two-component landscape on a 208×208 grid produced by a stochastic cellular automaton after 1000 iterations with random initial conditions. When a 60×60 block of data is removed from the center, the plausibly realistic image on the right is generated after 5000 iterations of a stochastic cellular automaton with replacements chosen randomly from an eight-cell neighborhood.

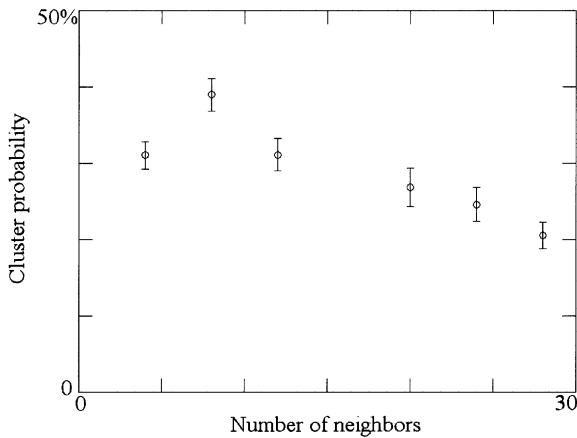


Fig. 2. Cluster probability produced by a two-component cellular automaton with replacements chosen randomly from different neighborhood sizes.



Fig. 3. The fraction of correctly identified cells in Fig. 1 is about 70% at the boundary and smoothly degrades to about 50% as the boundary recedes.

example, be chosen with a Gaussian or other radial probability distribution function.

Such multi-parameter probability kernels would permit fine-tuning the fit to reproduce the pattern even more realistically, for example using a spectrum of cluster probabilities in which more than four nearest neighbors are considered. Anisotropic patterns can be fit using an elliptical or rectangular neighborhood if desired.

Now suppose that the pattern in the left of Fig. 1 has a large block of 60×60 cells with missing data as indicated in the center plot of Fig. 1. Applying the cellular automaton model with replacements chosen from the eight nearest neighbors (Moore neighborhood) produces the pattern on the right in Fig. 1 after 5000 iterations. The number of iterations is not critical but should be sufficient to ensure that all the missing data cells are replaced and the resulting pattern has organized to the desired cluster probability. The block fills in gradually from the boundaries with a different but morphologically similar pattern to the one on the left of Fig. 1 without disturbing the surrounding region, which is assumed to contain good data.

The reconstructed data obviously do not duplicate the actual missing data, but they have the same general characteristics with the same probability distribution as the good data along the boundary. They blend into the good data in a seamless manner as the boundary is not discernible in the reconstructed image. Accurate prediction is not a goal of the method, but Fig. 3 shows an accuracy of about 70% near the boundary that degrades smoothly toward 50% as the boundary recedes as would be expected by pure chance. Fig. 3 is an average of many realizations of the predicted data as the cellular automaton evolves over 5000 iterations.

To demonstrate that the method works for more complex real landscape data, the left image in Fig. 4

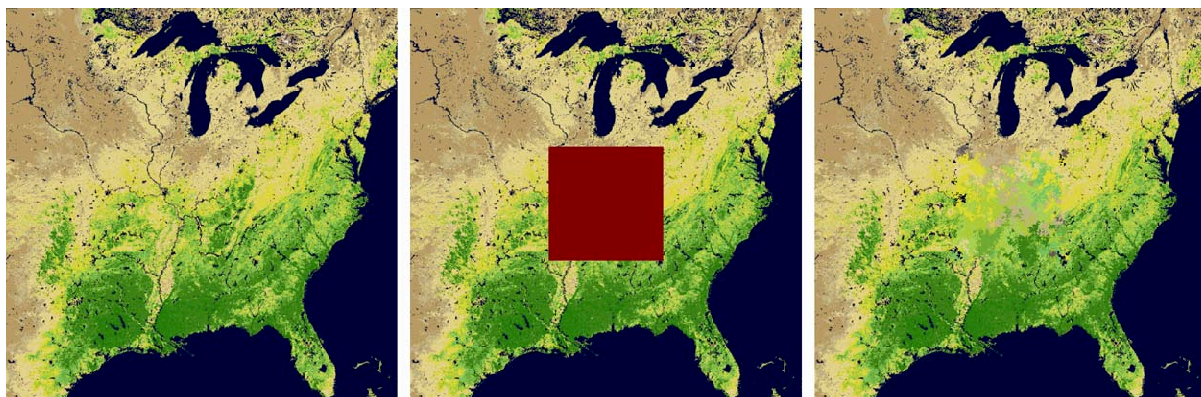


Fig. 4. The eight-level satellite data on a 548×548 grid of leaf area index over the Eastern United States on the left (courtesy of Steven Running, MODIS Land Group Member, University of Montana) is assumed to have a 160×160 block of data missing from the center and is reconstructed with 1000 iterations of a stochastic cellular automaton, producing the image on the right.

shows leaf area index (fraction of the surface area covered by green foliage) over the Eastern United States acquired by the Moderate-resolution Imaging Spectroradiometer (MODIS) during the period March 24–April 8, 2000 from the Terra satellite [20]. The image is

digitized with 548×548 pixels in eight colors (plus dark blue for water). In the center of Fig. 4, a 160×160 block of data is removed (perhaps to simulate a region obscured by clouds), and is reconstructed using 1000 asynchronous iterations of the stochastic cellular

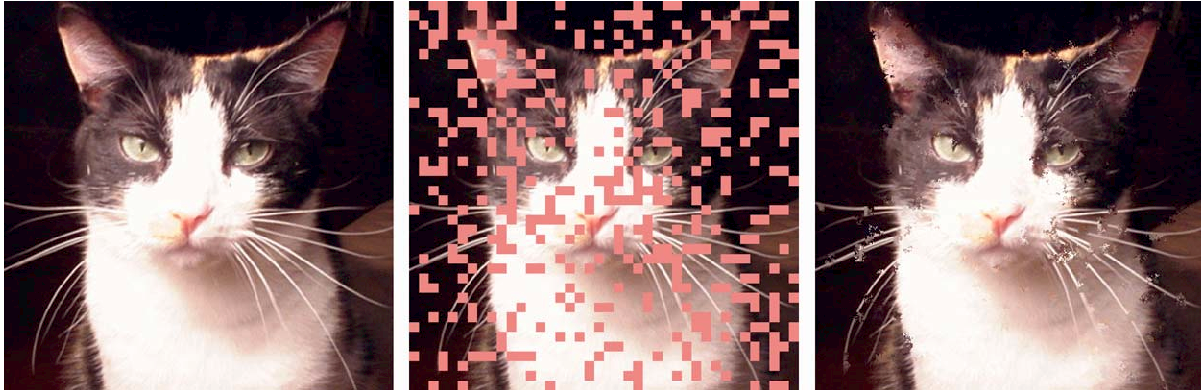


Fig. 5. The 256-color dithered image of a cat on the left is assumed to have 400 random blocks of 10×10 pixels removed and then replaced after 1000 iterations of a stochastic cellular automaton, producing the image on the right.

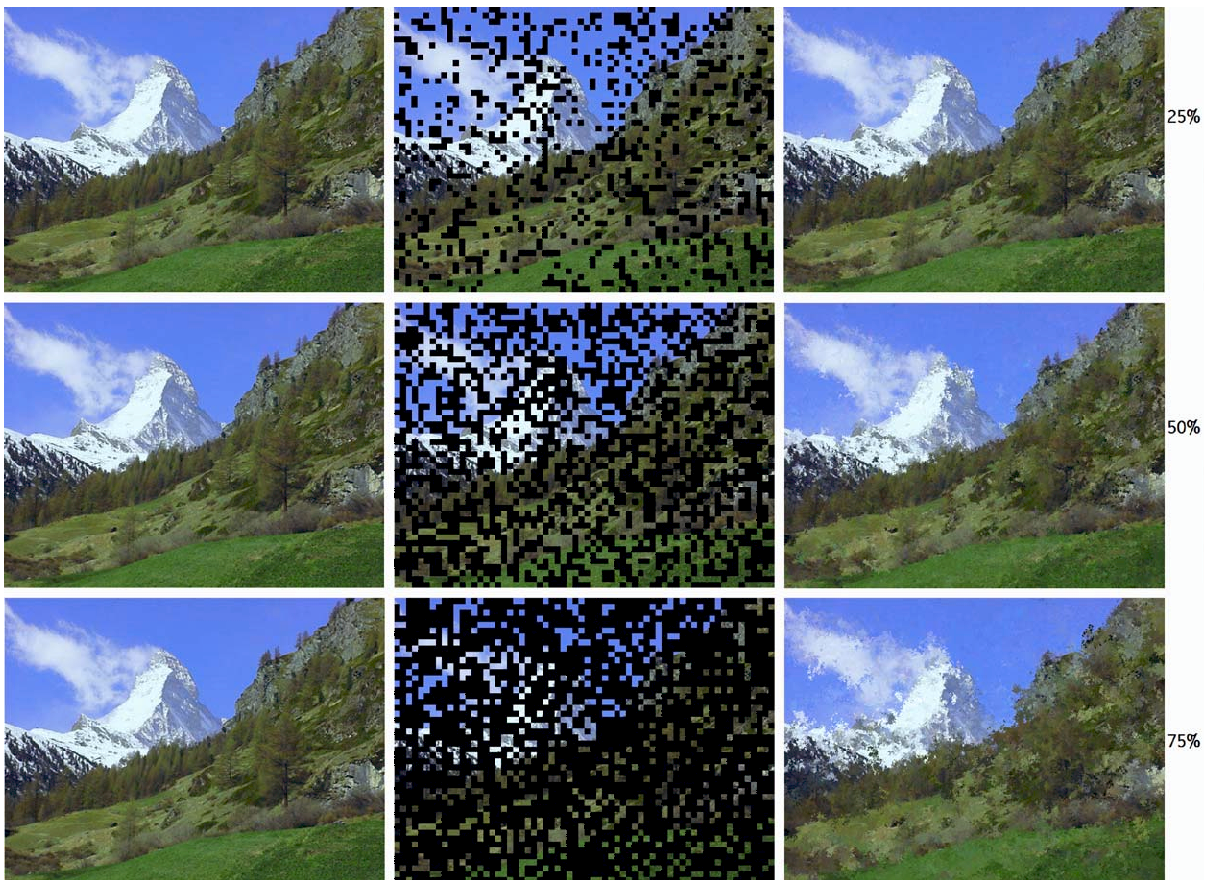


Fig. 6. The 256-color dithered image of the Matterhorn on the left is assumed to have 25%, 50%, and 75% of the pixels removed in random blocks of 10×10 pixels and then replaced after 1000 iterations of a stochastic cellular automaton, producing the images on the right.

automaton with replacements chosen randomly from an eight-cell neighborhood, producing the image on the right of Fig. 4. The reconstructed image is plausibly realistic with structure that resembles the actual structure and no discernable discontinuity at the boundary of the replaced region.

4. Application to digital photographs

For a more demanding and illustrative test of the method, Fig. 5 shows on the left a dithered image of a cat digitized at 400×400 pixels with 256 colors. The middle image shows the pattern after 400 blocks of 10×10 pixels have been removed from random positions within the image (25% of the pixels). On the right is the result of filling in the blocks using 1000 asynchronous iterations of the cellular automaton with replacements chosen randomly from an eight-cell neighborhood. The reconstructed image shows artifacts around the whiskers and the smooth edge of the ears, but in the regions of the face where the boundary between colors is more nearly fractal, the resulting image appears quite realistic. The dithering of the image to simulate more than 256 colors suggests that the method would also work well for half-tone photographs.

The method is especially suited for fractal landscapes that lack smooth edges such as the photograph of the Matterhorn digitized at 640×480 pixels with 256 colors in the left of Fig. 6. The three rows illustrate the effect of removing 25%, 50%, and 75% of the pixels, respectively, in random blocks of 10×10 pixels each and then reconstructed using 1000 asynchronous iterations of the cellular automaton with replacements chosen randomly from an eight-cell neighborhood. The image quality degrades gracefully, and even the worst case resembles an impressionist painting of the scene.

5. Conclusions

In summary, the proposed method uses a stochastic cellular automaton to generate fictitious but plausibly realistic fractal data in regions of space where actual data are missing or corrupted as evidenced by measures such as the cluster probability, fractal dimension, and algorithmic complexity. It does so by a simple method whose only parameter is easily determined from whatever good data exist. The method is mainly cosmetic and can be applied to any pattern or image whose underlying structure is a random fractal.

Acknowledgements

I am grateful to Janine Bolliger for suggesting this application, to David Mladenoff for supplying the

landscape data that motivated the idea, and to Cliff Pickover for advice and discussion.

References

- [1] Stewart LO. Public land surveys—history, instructions, methods. Ames, IA: Collegiate Press; 1935.
- [2] Mladenoff DJ, Dahir SE, Nordheim EV, Schulte LA, Guntenspergen GG. Narrowing historical uncertainty: probabilistic classification of ambiguously identified tree species in historical forest survey data. *Ecosystems* 2002;5:539–53.
- [3] Wang D, Vagnucci A, Li C. Digital image enhancement: a survey. *Computer Vision, Graphics, and Image Processing* 1983;24:363–81.
- [4] Stark H. Image recovery: theory and application. Orlando, FL: Academic Press; 1987.
- [5] Jain AK. Fundamentals of digital image processing. Englewood Cliffs, NJ: Prentice-Hall; 1989.
- [6] Schalkoff RJ. Digital image processing and computer vision. New York: Wiley; 1989.
- [7] Lewis R. Practical digital image processing. New York: Ellis Horwood; 1990.
- [8] Lim JS. Two-dimensional signal and image processing. Englewood Cliffs, NJ: Prentice-Hall; 1990.
- [9] Pratt WK. Digital image processing, 2nd ed. New York: Wiley; 1991.
- [10] Castleman R. Digital image processing. Englewood Cliffs, NJ: Prentice-Hall; 1996.
- [11] Russ JC. The image processing handbook, 3rd ed. Boca Raton, FL: CRC Press; 1999.
- [12] Gonzalez RC, Woods RE. Digital image processing, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall; 2002.
- [13] Hernandez G, Herrmann J, Goles E. External automata for image sharpening. *International Journal of Modern Physics C* 1994;5:923–31.
- [14] Hernandez G, Herrmann J. Cellular automata for elementary image enhancement. *Graphical Models and Image Processing* 1996;58:82–9.
- [15] Yang T. Cellular image processing. New York: Nova Science Publishers; 2001.
- [16] Bolliger J, Sprott JC, Mladenoff DJ. Self-organized criticality and complexity in historical landscape patterns. *Oikos* 2003;100:541–53.
- [17] Holley R, Liggett TM. Ergodic theorems for weakly interacting particle systems and the voter model. *Annals of Probability* 1975;3:643–63.
- [18] Sprott JC, Bolliger J, Mladenoff DJ. Self-organized criticality in forest-landscape evolution. *Physics Letters A* 2002;297:267–71.
- [19] Bak P. How nature works: the science of self-organized criticality. New York: Copernicus; 1996.
- [20] Myneni RB, Hoffman S, Glassy J, Zhang Y, Votava P, Nemani R, Running S, Privette JL, et al. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment* 2002;83:214–31.